

AN 10/613,261

Page 6

REMARKS

Claims 1-9 and 11-15 have been rejected under 35 U.S.C. §102(e) as being anticipated by the Taira U.S. Patent Application Publication No. 2003/0105638. The Office Action essentially takes the position that the Taira publication discloses a method for generating training data that can be used to normalize abbreviations. In support of this position the Office Action refers to an example in paragraph 0088, where the abbreviation is said to be "5 cm" and the expansion of the abbreviation is the word "large" in paragraph 0090. This action is said to create a logical relation having the expansion word "large" as one of three possible logical relations for the meaning of "5 cm". The Office Action then concludes that the result is training data used by the knowledge base to correctly determine whether the expansion is related to the phrase "5 cm".

The applicant respectfully disagrees with the position asserted in the Office Action. The method recited by claims 1-15 generates training data that can be used to normalize abbreviations in text. The term "abbreviations" is used in the claims to mean both "abbreviation" and "acronym" since these two words are interchangeable for the purposes of the invention. This ability to generate training data for abbreviations is useful in many natural language processing applications. As an example, large bodies of medical text containing substantial information have been created and maintained over significant periods of time. The information contained in this medical text offers substantial opportunities for advancement in the medical field. However, numerous and also highly ambiguous abbreviations are commonly used in this text. In order to make effective use of the text it is necessary to effectively normalize these abbreviations, i.e., to bring them to a common form for subsequent processing.

The Taira publication neither teaches nor suggests an abbreviation normalization method having these advantages. Rather, this publication teaches a method for generating training data for indicating logical relation instances for particular logical relation classes within a sentence. In a specific example referred to in the Office Action, the logical relation is between the "has size" classifier and the mass of "5 cm." Contrary to the position asserted in the Office Action, this reference does not describe the generation of training data for

AN 10/613,261

Page 7

normalizing the abbreviation "cm". The rejection of claims 1-15 should be withdrawn for this reason alone.

Another important difference between the claimed invention and the method shown in the Taira publication involves the level of automation required to generate the training data. As noted above, an important application of the invention is medical text processing. This text is extremely voluminous and replete with many often ambiguous abbreviations. The method of claims 1-15 is particularly well suited for this application because it generates the training data in a substantially automated fashion. To more particularly point out and distinctly claim this aspect of the invention, claim 1 has been amended to characterize the method as including the step of processing the corpus of text to identify the expansions in the text, and processing the corpus of text to generate the context information. The training data is then stored as a function of the context information generated by these processing steps.

Claims 13-15 are similar in that they recite a method for electronically generating the feature vectors. In particular, each stored feature vector includes context information generated for an associated expansion.

This method and its advantages are neither taught nor suggested by the Taira publication. The publication describes the training data creation step as one of supervised learning. Samples are submitted to the system for processing. However, they are then compared to results manually processed by a human expert (paragraph 0086). The human expert working to train the system must indicate all logical relation instances for a particular logical relation class within a sentence (paragraph 0087). For these additional reasons, rejection of claims 1-15 should be withdrawn.

The objection to claim 10 is noted, and the statement that it would be allowable if rewritten in independent form is appreciated. However, in view of the amendments and remarks presented above, it is believed that claim 10 is allowable in its present form.

BEST AVAILABLE COPY

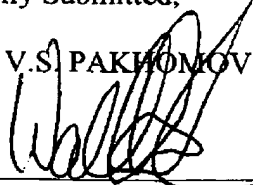
AN 10/613,261
Page 8

In conclusion, all claims 1-15 are in condition for allowance. Notice to that effect is respectfully requested.

Respectfully Submitted,

SERGEY V.S. PAKHOMOV

By:


Walter C. Linder, #31,707
FAEGRE & BENSON LLP
2200 Wells Fargo Center
90 South Seventh Street
Minneapolis, MN 55402-3901
612/766-8801

Dated: April 4, 2005

M2:20704368.01